1. Brad Eland [bradeland@gmail.com](mailto:%20bradeland@gmail.com) Phone # 801-903-9045

The title of my project is web scraping Linkedin   
Currently there is a person in Regis having to painstakingly search 1x1 to gather information/find previous MSDS students. I’m going to provide them with an automated solution.

This is the strict process of data engineering, not necessarily data science other than you need data to process things in data science. I do know that in the data engineering world this is a VERY beneficial skill to have. It does translate to being able to help in data science collection. It makes it so you can gather data (if not readily available) for data science use.

I expect to scrape the data from Linkedin based off of a list of people provided. So, in theory, it could be as big or small as is needed.

I won’t be analyzing data other than specifying whether or not I’m able to produce results for each of the records.

One of the biggest things I found from preliminary analysis of the business problem is the fact that there may be limits or ways Linkedin allows for scraping/bot detection. One thing I thought of was possibly creating a sleep or alternating sleeps when I have the process scraping so that “hopefully” the algorithm within Linkedin won’t block an account. I’m going to create a temporary or use another account to get the credentials. Just don’t want to get blocked/banned 😊

Since it was a minute settling on the project I still know I can get caught up in time to have a useful end product for Sarah within the Regis sphere.

Week 3 – Connect to selenium and get a new account within Linkedin. Begin uploading the data from .csv to python. I know I can path to a folder on my computer so whoever takes it/uses it in the future will just have to path to wherever on their machine. I’d be happy to assist anybody who may need help doing this.

Week 4 – Start processing records and returning what the end user would like returned based on business requirements from the meeting I have set 4/2 at 930 AM MST.

Week 5 – Connect an email service to python so I can send a .csv file to the end user(s). I will find out from the meeting in week 4 what platform is used within Regis and can try to set-up with my own credentials so somebody can plug in and do it.

Week 6 – Make sure that end users are getting their expected results via a .csv file to their respective emails.

Week 7 – Project wrap-up/create my presentation.